

AMAZON SALES ANALYSIS



-Madhugandha Pandav



Objective

- Gain actionable insights into Amazon sales data across three different branches of Amazon, respectively located in Mandalay, Yangon and Napyitaw by identifying and understanding the key factors influencing performance.
- This analysis aims to optimize strategies, enhance, profitability, and drive efficient decision making
- The data contains 17 columns and 1000 rows
- Approach Used:
- Data Wrangling: This is the first step where inspection of data is done to make sure NULL values and missing values are detected and data replacement methods are used to replace missing or NULL values.
- Feature Engineering: This will help us generate some new columns from existing ones.



Question-1

What is the count of distinct cities in the dataset?

```
select count(distinct city) as cities from `amazon (1)`;
```

	cities
▶	3



Question-2

For each branch, what is the corresponding city?

```
select distinct branch, city from `amazon (1)`;
```

	branch	city
▶	A	Yangon
	C	Naypyitaw
	B	Mandalay



Question-3

What is the count of distinct product lines in the dataset?

```
Select count( distinct `Product line`) as Prouduct_line  
FROM `amazon (1)`;
```

	Prouduct_line
▶	6



Question-4

Which payment method occurs most frequently?

```
select max(payment) as payment_method from `amazon (1)`;
```

	payment_method
▶	Ewallet



Question-5

Which product line has the highest sales?

```
select `Product line`, -- selecting the product line  
count(`Invoice ID`) as sales_count -- counting the number of 'invoice id' by aliasing  
from `amazon (1)`  
group by `product line` -- grouping the results by product line  
order by sales_count desc -- ordering the results by sales_count(no. of invoice id) in des  
limit 1;-- limiting the result to only the top (highest) sales count
```

Product line	sales_count
Fashion accessories	178



Question-6

How much revenue is generated each month?

```
select month_name, count(`invoice id`) as monthly_revenue  
from `amazon (1)`  
group by `month_name`;
```

month_name	monthly_revenue
Jan	352
Mar	345
Feb	303



Question-7

In which month did the cost of goods sold reach its peak?

```
select month_name, sum(cogs) as total_cogs
from `amazon (1)`
group by month_name -- grouping the result by month
order by total_cogs desc -- ordering the results by total_cogs(total cogs) in desc order
limit 1; -- limiting the result to only the top(highest) total_cogs
```

	month_name	total_cogs
▶	Jan	110754.160000000002

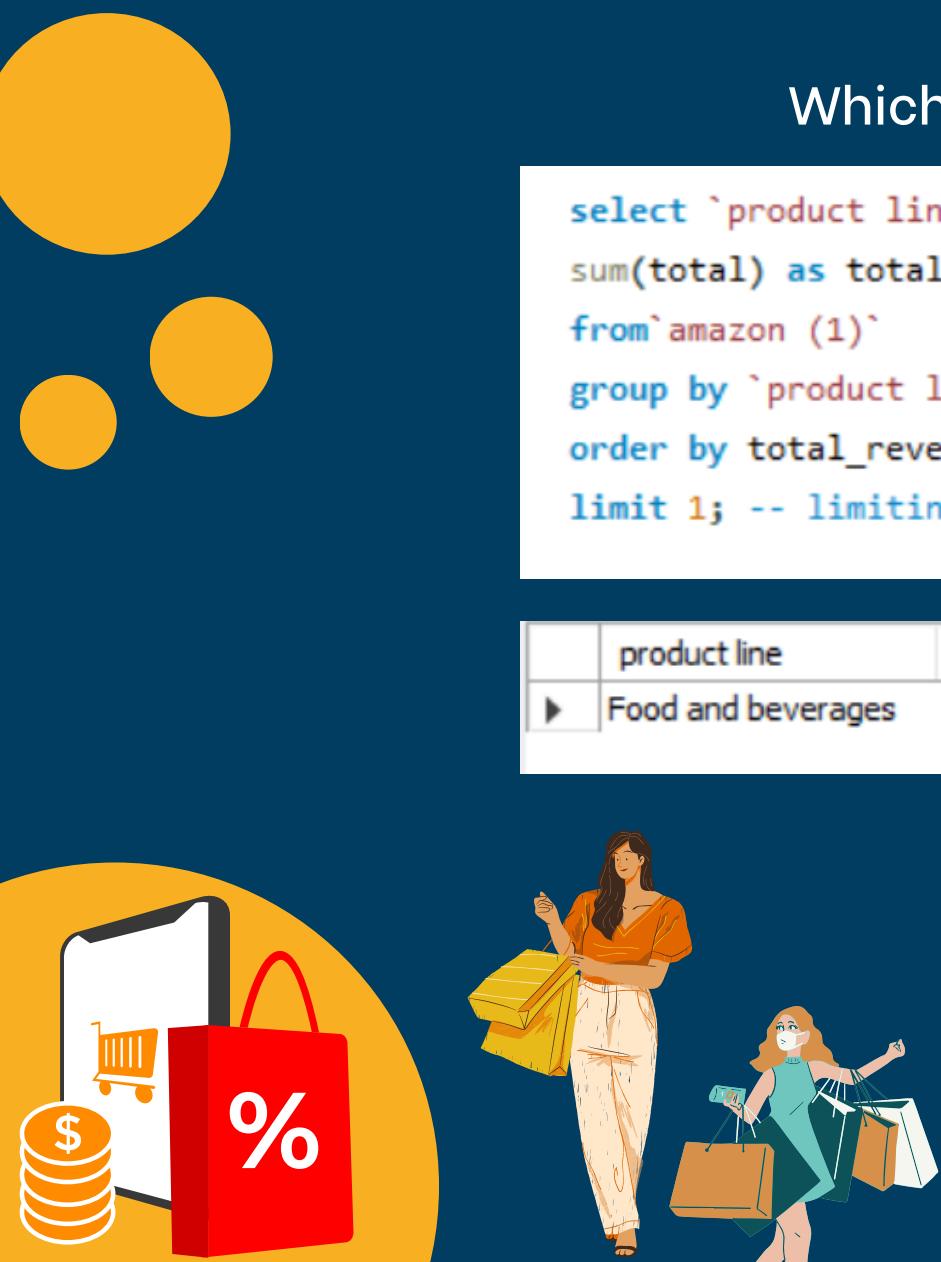


Question-8

Which product line generated the highest revenue?

```
select `product line`,
sum(total) as total_revenue -- calculating total revenue for each product line
from`amazon (1)`
group by `product line`
order by total_revenue desc
limit 1; -- limiting the result to only top (highest) total_revenue
```

	product line	total_revenue
▶	Food and beverages	56144.844000000005



Question-9

In which city was the highest revenue recorded?

```
select city, sum(total) as high_revenue  
from `amazon (1)`  
group by city  
order by high_revenue desc;
```



	city	high_revenue
▶	Naypyitaw	110568.70649999994
	Yangon	106200.3705000001
	Mandalay	106197.67199999996

Question-10

Which product line incurred the highest Value Added Tax?

```
select `product line`,  
sum(`tax 5%`) as total_vat_amount --  
from `amazon (1)`  
group by `product line`  
order by total_vat_amount desc  
limit 1;
```

	product line	total_vat_amount
▶	Food and beverages	2673.563999999994



Question-11

For each product line, add a column indicating “Good” if its sales are above average, otherwise “Bad.”

```
select `product line`,  
Case  
when `gross income` > (select avg(`gross income`) from `amazon (1)` ) then "good"  
else "bad"  
end as sales_performance  
from `amazon (1)`;
```

	product line	sales_performance
▶	Health and beauty	good
	Electronic accessories	bad
	Home and lifestyle	good
	Health and beauty	good
	Sports and travel	good
	Electronic accessories	good
	Electronic accessories	good

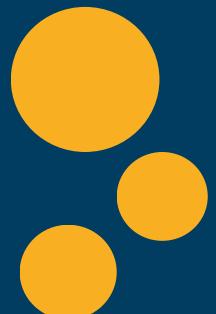


Question-12

Identify the branch that exceeded the average number of products sold.
select distinct branch from `amazon (1)`

```
select distinct branch from `amazon (1)`  
where quantity > (  
    select avg(quantity)  
    from `amazon (1)`);  
-- All three branches have exceeded the avg number of product
```

	branch
▶	A
	C
	B



Question-13

Which product line is most frequently associated with each gender?

```
With ranked_product_lines as
  ( select gender, `product line`, count(*) as product_line_count,
    rank() over(partition by gender order by count(*) desc) as rank_num
  from `amazon (1)`
  group by gender, `product line`)
  select gender, `product line`, product_line_count
  from ranked_product_lines
  where rank_num = 1;
```

	gender	product line	product_line_count
▶	Female	Fashion accessories	96
	Male	Health and beauty	88

Question-14

Calculate the average rating for each product line.

```
select `product line`, avg(rating) as avg_rating  
from `amazon (1)`  
group by `product line`;
```



	product line	avg_rating
▶	Health and beauty	7.003289473684212
	Electronic accessories	6.92470588235294
	Home and lifestyle	6.8375
	Sports and travel	6.916265060240964
	Food and beverages	7.113218390804598
	Fashion accessories	7.029213483146067

Question-15

Count the sales occurrences for each time of day on every weekday.

```
select day_names, time_of_day, count(*) as sale_occur  
from `amazon (1)`  
group by day_names, time_of_day  
order by day_names, sale_occur desc ;
```

	day_names	time_of_day	sale_occur
▶	Fri	afternoon	74
	Fri	evening	36
	Fri	morning	29
	Mon	afternoon	75
	Mon	evening	29
	Mon	morning	21
	Sat	afternoon	81



Question-16

Identify the customer type contributing the highest revenue.

```
select `customer type`, sum(total) as revenue -- calcaulating the total revenue for each customer type
from `amazon (1)`
group by `customer type`
order by revenue desc;
```

	customer type	revenue
▶	Member	164223.44400000002
	Normal	158743.30500000005



Question-17

Determine the city with the highest VAT percentage.

```
select city,
       sum(`Tax 5%`) as high_vat, sum(total) as total_rev,
       (sum(`tax 5%`)/sum(total))*100 AS vat_percent -- calculating the vat percenage
  from `amazon (1)`
 group by city
 order by vat_percent desc;
```

	city	high_vat	total_rev	vat_percent
▶	Naypyitaw	5265.176500000002	110568.70649999994	4.761904761904766
	Mandalay	5057.032000000003	106197.67199999996	4.761904761904766
	Yangon	5057.160500000002	106200.3705000001	4.761904761904759



Question-18

Identify the customer type with the highest VAT payments.

```
select `customer type`,  
sum(`tax 5%`) as high_vat  
from `amazon (1)`  
group by `customer type`  
order by high_vat desc;
```

	customer type	high_vat
▶	Member	7820.1640000000002
	Normal	7559.2050000000003



Question-19

What is the count of distinct customer types in the database?

```
select count(distinct `customer type`) as customer_type_count  
from `amazon (1)`;
```

	customer_type_count
▶	2



Question-20

What is the count of distinct payment methods in the dataset?

```
select count(distinct payment) as payment_method_count  
from `amazon (1)`;
```

	payment_method_count
▶	3



Question-21

Which customer type occurs most frequently?

```
select `customer type`, count(*) as most_freq  
from `amazon (1)`  
group by `Customer type`;
```

	customer type	most_freq
▶	Member	501
	Normal	499



Question-22

Identify the customer type with the highest purchase frequency.

```
select `customer type`, sum(quantity) as high_pur_freq  
from `amazon (1)`  
group by `Customer type`  
order by high_pur_freq  
limit 1;
```

	customer type	high_pur_freq
▶	Normal	2725



Question-23

Determine the predominant gender among customers.

```
select gender, count(*) customer_count  
from `amazon (1)`  
group by gender  
order by customer_count desc  
limit 1;
```

	gender	customer_count
▶	Female	501



Question-24

Examine the distribution of genders within each branch.

```
select branch, gender, count(gender) as gender_dist  
from `amazon (1)`  
group by branch, gender  
order by branch, gender_dist desc;
```

	branch	gender	gender_dist
▶	A	Male	179
	A	Female	161
	B	Male	170
	B	Female	162
	C	Female	178
	C	Male	150



Question-25

Identify the time of day when customers provide the most ratings.

```
select time_of_day, count(rating) as rating_count  
from `amazon (1)`  
group by time_of_day  
order by rating_count desc;
```

	time_of_day	rating_count
▶	afternoon	528
	evening	281
	morning	191



Question-26

Determine the time of day with the highest customer ratings for each branch.

```
select branch, time_of_day, count(rating) as rating_count
from `amazon (1)`
group by branch, time_of_day
order by branch desc;
```

	branch	time_of_day	rating_count
▶	C	afternoon	181
C		evening	88
C		morning	59
B		afternoon	162
B		evening	111
B		morning	59
▲		afternoon	185



Question-27

Identify the day of the week with the highest average ratings.

```
select day_names, avg(rating) as avg_rating
from `amazon (1)`
group by day_names
order by avg_rating desc
limit 1;
```

	day_names	avg_rating
▶	Mon	7.153599999999999



AMAZON SALES ANALYSIS

Question-28

Determine the day of the week with the highest average ratings for each branch.



```
WITH branch_high_rating AS
  (select branch, day_names, avg(rating) as avg_rating,
   rank() over(partition by branch order by avg(rating) desc) as rank_num
    from `amazon` (1)
   group by branch, day_names)
  select branch, day_names, avg_rating
  from branch_high_rating
  where rank_num = 1;
```

	branch	day_names	avg_rating
▶	A	Fri	7.3119999999999985
	B	Mon	7.335897435897434
	C	Fri	7.278947368421051





PRODUCT ANALYSIS

□ 6 Product Lines are present:

- ❖ Health and beauty
- ❖ Electronic accessories
- ❖ Home and Lifestyle
- ❖ Sports and Travel
- ❖ Food and beverages
- ❖ Fashion accessories

□ **Food and beverages** is generating **highest revenue** of around 56145 and incurred **highest VAT** of around 2673.56 while **Electronic accessories** has recorded **highest sales**.

□ **Health and beauty** is generating **lowest revenue**, **lowest sales** as well as incurred **lowest VAT** among all product lines.



AMAZON SALES ANALYSIS



SALES ANALYSIS

- ❑ **January** has generated **highest revenue** of around **116292**.
- ❑ **Branch A** has exceeded the average sales and sold **1859 total products**.
- ❑ **Naypyitaw** has generated **highest revenue** and cash is most used payment method in this city.
- ❑ **Afternoon** time has recorded the **highest sales** while Morning has recorded the lowest sales.
- ❑ There are **THREE payment methods** viz. Cash, Credit Card, Ewallet. Out of these three **EWALLET** is **frequently used** by customers.



AMAZON SALES ANALYSIS



CUSTOMER ANALYSIS

- There **2 customer types** viz. **member and normal**. **Member** customer type has contributed **more** in terms of **revenue** as well as highest **purchasing frequency**.
- **Females** have contributed **more to the revenue**, although difference between contribution of males and females is not much.
- **Females** are more inclined towards the product line **Fashion accessories** while **males** are associated with **Health and beauty**.



AMAZON SALES ANALYSIS



RECOMMENDATIONS

- **Health and Beauty product** is least performing on all matrices and needs attention. **Proper plan** needs to be span out to enhance this product line.
- **Diverse options** for customers and more **effective and targeting strategies** can be followed **during January** as this month has recorded highest sales and generated highest revenue.
- **Afternoon** is being observed as **peak sales hours**, accordingly businesses can prioritize launching **new products or campaigns** during this hour.
- **Member type** customers have **contributed more**, this might be due incentives or offers hence **proper plan** can be chalked out to **give memberships** to more customers as this **ultimately reduce Customer Acquisition Cost** and **generate more revenue**.



AMAZON SALES ANALYSIS



Thank you!!



AMAZON SALES ANALYSIS